Prandi Davide (Orcid ID: 0000-0001-9885-6074)
Rubin Mark (Orcid ID: 0000-0002-8321-9950)

24
21-533.R2 Poulsom-EDITED USA English ORIGINAL ARTICLE

Comparative genomics of primary prostate cancer and paired metastases: insights from 12 molecular case studies

Joanna Cyrta[1,2,3] ‡, Davide Prandi[4] §, Arshi Arora[5], Daniel H. Hovelson[6], Andrea Sboner[2,7], Antonio Rodriguez[3,8], Tarcisio Fedrizzi[5], Himisha Beltran[9,10], Dan R. Robinson[11], Anurandha Gopalan[12], Lawrence True[13], Peter S. Nelson[14], Brian D. Robinson[1,2], Juan Miguel Mosquera[1,2], Scott A. Tomlins[11] †, Ronglai Shen[5] †, Francesca Demichelis[2,4] †, Mark A. Rubin[1,2,3] † *

[1]Department of Pathology and Laboratory Medicine, Weill Cornell Medicine, New York, NY, USA

[2]Englander Institute for Precision Medicine, Weill Cornell Medicine, New York, NY, USA

[3]Department for BioMedical Research, University of Bern, Bern, Switzerland

‡ Current affiliation: Department of Pathology, Institut Curie, Paris, France

[4]Department of Cellular, Computational and Integrative Biology, University of Trento, Trento, Italy

§ Current affiliation: Fondazione Bruno Kessler, Povo, Italy

[5]Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, NY, USA

[6]Center for Computational Medicine and Bioinformatics, Univ. Michigan, Ann Arbor, MA, USA

[7]HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine, Weill Cornell Medicine, New York, NY, USA

[8]Institute of Pathology, University of Bern, Bern, Switzerland

[9]Department of Medicine, Division of Medical Oncology, Weill Cornell Medicine, New York, NY, USA

[10]Department of Medical Oncology, Dana Farber Cancer Institute, Boston MA, USA

[11]Department of Pathology, University of Michigan, Ann Arbor MI, USA

[12]Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY, USA

[13]Department of Pathology, Univ. of Washington, Seattle, WA USA

[14]Fred Hutchinson Cancer Research Center, Seattle, WA, USA

† These authors contributed equally to the work

*Correspondence to: MA Rubin, University of Bern, Department for BioMedical Research, Murtenstrasse 35, CH-3008 Bern, Switzerland. E-mail: mark.rubin@dbmr.unibe.ch

Funding statement: This study was supported by the 2015-2017 Movember Foundation-PCF Challenge Award. It has also received funding from the NIH/NCI WCM SPORE in Prostate Cancer P50-CA211024 (H.B., F.D., M.A.R., J.M.M., B.D.R.), the Nuovo-Soldati Foundation (J.C.), the National Science Foundation Project grants number R37 CA241486 (H.B.), and R01 CA125612 (F.D.), the Pacific Northwest Cancer Spore Award P50CA097186 and the CDMRP Award W81XWH-18-1-0406 (P.N.), and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 648670) (F.D.).

Conflict of interest disclosure: The University of Michigan and Harvard have a patent in the field of diagnostics and therapeutics for ETS gene fusion prostate cancer; ST, FD and MAR are named as co-inventors. The diagnostic field of use was licensed to Hologic/Gen-Probe (who sublicensed some rights to Ventana/Roche) and is licensed to LynxDx. Cornell University has a patent on SPOP prostate cancer mutations in the field of diagnostics and therapeutics; MAR is a co-inventor. MAR serves on the Scientific Advisory Board of Neogenomics. ST is a co-inventor on a patent issued to Strata Oncology related to MSI determination and checkpoint inhibitor benefit. ST has the following other conflicts of interest: Employment (Strata Oncology); Leadership (Strata Oncology), Stock and Other Ownership Interests (Strata Oncology, Javelin Oncology); Consulting or Advisory Roles (Janssen, Astellas Medivation, Strata Oncology);

Research Funding (Astellas Medivation); Travel, accommodation, and expenses (Strata Oncology, Genzyme).

Running title: Genomics of primary prostate cancer and matched metastatic disease.

Word count: 3,655

Abstract

Primary prostate cancer (PCa) can show marked molecular heterogeneity. However, systematic analyses comparing primary PCa and matched metastases in individual patients are lacking. We aimed to address the molecular aspects of metastatic progression while accounting for heterogeneity of primary PCa.

In this pilot study, we collected 12 radical prostatectomy (RP) specimens from men who subsequently developed metastatic castration-resistant prostate cancer (mCRPC).

We used histomorphology (Gleason grade, focus size, stage) and immunohistochemistry (IHC) (ERG and p53) to identify independent tumors and/or distinct subclones of primary PCa. We then compared molecular profiles of these primary PCa areas to matched metastatic samples using whole exome sequencing (WES) and amplicon-based DNA and RNA sequencing.

Based on combined pathology and molecular analysis, seven (58%) RP specimens harbored monoclonal and topographically continuous disease, albeit with some degree of intra-tumor heterogeneity; four (33%) specimens showed true multifocal disease; and one displayed monoclonal disease with discontinuous topography. Early (truncal) events in primary PCa included *SPOP* p.F133V (one patient), *BRAF* p.K601E (one patient), and *TMPRSS2*:ETS rearrangements (eight patients). Activating *AR* alterations were seen in nine (75%) mCRPC patients, but not in matched primary PCa. Hotspot *TP53* mutations, found in metastases from three patients, were readily present in matched primary disease. Alterations in genes encoding epigenetic modifiers were observed in several patients (either shared between primary foci and metastases or in metastatic samples only).

WES-based phylogenetic reconstruction and/or clonality scores were consistent with the index focus designated by pathology review in six out of nine (67%) cases. The three instances of discordance pertained to monoclonal, topographically continuous tumors, which would have been considered as unique disease in routine practice.

Overall, our results emphasize pathologic and molecular heterogeneity of primary PCa, and suggest that comprehensive IHC-assisted pathology review and genomic analysis are highly concordant in nominating the “index” primary PCa area.

Keywords: castration-resistant; CRPC; genomics; heterogeneity; metastasis; multifocal; NEPC; pathology; prostate cancer; transcriptomics.

Introduction

Primary prostate cancer (PCa) is a multifocal disease in up to 80% of PCa patients [1,2]. In addition, it has been shown to display marked inter- and intra-tumor heterogeneity at the genomic, transcriptomic and DNA methylation levels [3–8]. Conversely, metastatic castration-resistant prostate cancer (mCRPC) appears to be of clonal origin, even though it can acquire molecular heterogeneity through subclonal evolution [4,7,9]. There is a critical knowledge gap regarding the molecular mediators of PCa progression to mCRPC in individual patients. While pathology criteria associated with aggressive disease, e.g. Gleason grade or focus size, are typically used to nominate the dominant (index) focus in radical prostatectomy (RP) specimens, the validity of this approach has not been addressed by systematic studies, and has even been challenged by occasional case reports [10].

Due to the long time between primary therapy and development of metastatic disease, only a few studies have analyzed matched primary and metastatic samples. As part of two precision oncology trials, the CRPC500 [11,12] and the Weill Cornell Medicine (WCM) Precision Oncology cohort [13], we present a comprehensive pathology and genomic analysis of RP specimens from 12 men who subsequently developed mCRPC. We compare the genomics of multiple areas of primary PCa with matched metastatic disease, with the aim of nominating the index focus and identifying the histopathological and molecular criteria that could be associated with metastatic outcome.

Materials and methods

Ethics approval statement.

The study included patients enrolled in the Weill Cornell Medicine precision cancer care program and/or in the Stand-Up-To-Cancer (SU2C) CRPC500 study with appropriate IRB approval and written informed consent.

Patient consent statement

Written informed consent for molecular studies was previously obtained from all patients as part of the Weill Cornell Medicine precision cancer care program and/or in the Stand-Up-To-Cancer (SU2C) CRPC500 study.

Sample collection and pathology review

Medical records of patients with mCRPC enrolled in the WCM precision cancer care program and/or in the Stand-Up-To-Cancer (SU2C) CRPC500 study [12] were interrogated to identify patients who had previously undergone radical prostatectomy (RP). Only patients for whom complete RP pathology material (all H&E slides and blocks) could be retrieved were included. Three pathologists (JMM, BR and JC) jointly reviewed all H&E slides for each RP specimen. Immunohistochemistry (IHC) for ERG and p53 was performed on all tumor areas. Potentially distinct areas (i.e., either subclones of the same primary disease or truly independent tumors) were selected for sequencing based on topography (i.e., discontinuous nature of tumor areas), individual Gleason score, histomorphology, and/or IHC. The putative index focus was nominated based on the highest individual Gleason score (primary criterion) and focus size (secondary criterion). Seminal vesicle invasion and metastatic regional lymph nodes removed at time of RP were also included in this evaluation and by principle, considered as putative index disease. Pathology review of metastatic samples (frozen sections) was previously performed as part of the WCM precision cancer care and/or the SU2C CRPC500 study [12]. Both studies were performed with appropriate written patient consent and IRB approval.

Immunohistochemistry

Immunohistochemistry (IHC) was performed on sections of FFPE tissue using a Bond III automated immunostainer and the Bond Polymer Refine detection system (Leica Microsystems, Buffalo Grove, IL, USA), with the following antibodies and conditions (dilution, heat-mediated antigen retrieval solution, retrieval time): anti-ERG (clone EPR386, 1/100, H1, 30 min; Abcam, Cambridge, MA, USA); anti-p53 (clone DO-7, 1/100, H1, 20 min; Cell Signaling Technologies, Danvers MA, USA).

Whole-exome sequencing

WES was performed using previously validated WCM protocols [14]. In brief, DNA was extracted from macrodissected unstained slides of FFPE tissue cut at 10 µm (for each primary tumor focus); from cored frozen, OCT-embedded tissue (for metastases); and from peripheral blood lymphocytes (for germline control). DNA extraction was accomplished using Promega Maxwell 16 MDx (Promega, Madison, WI, USA). DNA quality was confirmed using a Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). Libraries were prepared using exome capture of 21,522 genes with the HaloPlex System (Agilent). Sequencing was performed on Illumina HiSeq 2500 (Illumina, San Diego, CA< USA) in 100 bp paired-end mode [14]. Reads were aligned to GRC37/hg19 reference using BWA [15] and processed accordingly to Whole Exome Sequencing Test for Cancer – ExaCT-1 - pipeline v0.9 [14].

To identify SNVs, we applied both MuTect [16] and SNVseeqer [17] to nominate putative aberrant genomic positions. Identified genomic positions were filtered requiring coverage of at least 10 reads, a read count of the alternative base of at least 3, and a minimum variant allelic fraction (VAF) of 5%. Filtered positions were inspected with ASEQ [18], which provides a read count for each of the four bases in tumor and matched normal samples. Genomic positions where the read count of the alternative base in normal matched sample was greater than 0 were considered not aberrant. The list of aberrant genes was divided into 3 tiers: tier1 containing genes in the EXaCT-2 WCM test, tier2 containing genes in COSMIC Cancer Gene Census, tier3 containing all remaining genes.

For metastatic samples for Patients 11 and 12, WES data from fresh-frozen tissue were obtained through the CRPC500 study and generated as described previously [11].

Targeted DNA and RNA sequencing

DNA and RNA extraction, amplicon-based next generation sequencing, and bioinformatics analysis were performed as previously described [5].

Phylogenetic reconstruction and clonality analysis

Based on SNV calls, primary and metastatic samples were analyzed for clonal relatedness through phylogenetic tree reconstruction and clonality score. The final list of aberrant genomic positions was composed of: i) mutations in genes in tier1 or tier2 called by either Mutect or SNVseeqer, and ii) mutations in genes in tier3 called by both Mutect and SNVseeqer. Maximum-parsimony trees were inferred using the phangorn package [19]. The germline sample was designated as the root. Primary and metastatic tumor samples were plotted as descendent nodes, with edge lengths indicating the number of alterations newly accumulated in descendant nodes. Samples sharing common mutations formed a node. Clonality scores were determined as previously described [20] by performing conditional test on sites with mutation in one or both tumors. Marginal probabilities at each locus were estimated via empirical relative frequencies from the somatic mutations data set from the publicly available TCGA PRAD study. [21].

Results

Pathology and clinical findings

Twelve patients were included in this pilot cohort (nine enrolled in WCM precision cancer care and three in the SU2C CRPC500 study [12]). The number and type of samples per patient, pathology results and IHC are summarized in **Figure 1** and in **supplementary material, Table S1.** We assigned individual Gleason scores to each discrete sample. Samples showed a range from Gleason score 6 (3+3) to 9 (5+4). Disease stage per patient ranged from pT2aN0 to pT3bN1. Partial neuroendocrine features (positive synaptophysin expression with maintained AR expression and an adenocarcinoma morphology) were noted in primary PCa from one patient (Patient 7) and in the metastasis of another patient (Patient 6). Neuroendocrine prostatic carcinoma (NEPC) with adequate morphology [22] and loss of AR expression was seen in the RP specimen from one patient (Patient 12) and in metastases from two patients (Patients 12 and 7). ERG expression was seen in all or a subset of samples in 8 (67%) patients.

The time from prostatectomy to metastatic sample collection (per sample) ranged from 15.9 to 230.5 months (mean, 47.1 months); the shortest time interval was measured for a patient with NEPC primary tumor. Time intervals for patients whose primary disease harbored at least one focus with an individual Gleason score of 9 were significantly shorter compared to other patients (mean, 46 months versus 115.5 months, p=0.046, unpaired two-tailed *t*-test). Metastatic sites included bone, lymph node, liver and brain. The available information about patient treatments is summarized in **supplementary material, Table S2**.

Initial pathology review of RP specimens, including IHC results for ERG and p53, but blinded to the genomic data, interpreted eight (67%) RP specimens as monoclonal disease with intra-tumor heterogeneity and four (33%) RP specimens as harboring true multiclonal disease (Patients 3, 6, 11 and 12). In three of the multiclonal cases, the additional tumors were only small-volume Gleason grade 6 foci (Patients 3, 6 and 11). In one case (Patient 12), as many as six potentially independent tumors were identified, including three Gleason grade 7 and three

Gleason grade 6 lesions (**Figure 2**). Using Gleason grade, tumor size and stage, one or multiple index tumor areas were nominated by pathology for all cases.

Genomic landscape of primary PCa and paired metastatic samples

WES was performed on 36 primary tumor areas from 10 patients (2-6 samples per patient) (**supplementary material, Tables S3–S5, Figures 3 and 4, supplementary material, Figures S1–S8**). Targeted DNA sequencing was performed on 63 primary samples from 12 patients (3–7 per patient) (**supplementary material, Tables S6,S7**). Results were compared to metastatic tumor samples (18 samples from 12 patients, 1-2 samples per patient) which previously underwent WES [12,13].

The most frequent genomic event found in the metastases, but not in the matching primary samples, were *AR* activating alterations, seen in 9 (75%) patients (amplification in eight cases, and activating point mutations with amplification in one case). The second most frequent alteration type found in metastases were *TP53* missense hotspot mutations and *MYC* amplifications, each identified in three (25%) patients. Importantly, in contrast to *AR* activating alterations, matched analysis confirmed that these *TP53* alterations were present in some (two cases) or all (one case) areas of primary PCa in each patient (**Figures 1–3**).

Comparative analysis of paired samples from individual patients identified some early ("truncal") events, present at high allelic frequencies in all primary samples and in the metastases. These included: a hotspot p.F133V *SPOP* mutation in Patient 2, and a *BRAF* p.K601E mutation in Patient 4. In addition, RNA sequencing identified *TMPRSS2:ERG* and *TMPRSS2:ETV1* fusion transcripts in 6 patients and in one patient, respectively. These fusion transcripts were detected across multiple samples and, when data were available, in the matched metastasis (except for Patient 12). The *TMPRSS2:ERG* status was consistent with ERG IHC in all but one sample. This one sample (Patient 6, T4) was ERG-negative on IHC and displayed a relatively low *TMPRSS2:ERG* signal on RNA-seq, likely due to contamination with cells from the nearby ERG-positive tumor focus T3 (**supplementary material, Figure S9**). The

time from RP to metastatic biopsy was not significantly different between patients with ERG-positive and ERG-negative primary disease (unpaired two-tailed *t*-test).

Several cases harbored alterations in genes encoding epigenetic modifiers. *KMT2D (MLL2/4)* c.7683A_fs in Patient 9 was shared by all primary samples, except for one area with a lower Gleason score, and by the two metastases (**Figure 4**). Other alterations included: *KMT2C (MLL3)* p.C1013* in Patient 2, shared by all primary samples and the metastasis (**supplementary material, Figure S2**); *KDM4A* (*JMJD2*) p.L696_fs in Patient 4, detected in the metastasis only (**supplementary material, Figure S4**); and p.1479_fs in *ARID1A,* encoding a member of the SWI/SNF complex, detected in both metastatic samples in Patient 1, but not in the primary samples (**supplementary material, Figure S1**).

BRCA2 germline mutations were found in two (16.7%) cases (Patients 1 and 8), one of which showed neuroendocrine features in the primary and metastatic tumors.

Assessing clonal relatedness between primary and metastatic samples

Phylogenetic reconstruction using WES SNV calls allowed to nominate an index focus/foci in seven (70%) of the 10 patients for whom WES on the primary samples was available, and clonality scores could be obtained for nine patients (90%) (**Figure 1, supplementary material, Table S1**). Both approaches showed highly concordant results.

In six of the nine cases (67%), the index focus nominated by phylogenetic reconstruction and/or by the highest clonality score was consistent with the one proposed by pathology review. Discordance was seen in three cases (Patients 2, 3 and 4), all of which were monoclonal tumors with continuous spread.

In addition, although phylogeny and clonality analyses were not available for Patient 12, concordance between pathology and genomics could be inferred owing to the presence of a shared *TP53* mutation (p.R248W) and the absence of a *TMPRSS2:ERG* fusion in the index focus and the metastases (**Figure 1**).

Gene expression analysis

Targeted RNA sequencing was performed on 58 primary and 5 metastatic samples (**Figure 5**). Unsupervised hierarchical clustering using gene expression levels from RNA-seq further supported molecular similarity between genomically similar PCa areas, especially in monoclonal primary tumors. Conversely, sample clustering results for Patients 3, 6, 11 and 12 were potentially consistent with multiclonality. Of note, mCRPC samples from four patients (two bone and two liver metastases) formed a distinct transcriptomic cluster, characterized by high expression of proliferation genes (**Figure 5**, box), consistent with results using this approach on a separate compendium of localized treatment-naive PCa and CRPC metastases [5].

Discussion

Genomic heterogeneity of PCa and different scenarios for PCa clonality

Primary PCa has previously been shown to represent a molecularly heterogeneous disease [3-6]. When considering clonality of primary PCa, four scenarios can be proposed, all of which were encountered in our pilot cohort (**Figure 6, supplementary material, Table S1**).

In one scenario, tumors are monoclonal (all areas have a common origin) and show topographically continuous spread, but still demonstrate intra-tumor heterogeneity due to emergence of distinct subclones [3,4]. This was the most common scenario in our study, seen in seven (58%) cases. It has previously been shown that evolutionary branching usually precedes metastatic seeding, and genomic heterogeneity of localized PCa has even been suggested as a potential marker of aggressive disease [3,7].

Another possibility is the presence of true multiclonal tumors with independent origins in the RP specimen [23-26]. Only four (33.3%) cases from our cohort showed true multiclonal disease. The rate of true multiclonal disease in our study was lower than the 50-80% previously reported in large RP series [1,2]. This could be due to selection bias, as we only included patients who subsequently developed metastases, thus selecting for aggressive primary disease, which may potentially override synchronous low-volume lesions or not leave the time for such lesions to develop. This is in keeping with a study by Wise et al., who found that a greater number of “lesser cancers” correlates with a smaller volume of the index cancer [2]. In most instances,

and in three (25%) of our patients, such additional foci are low-grade and low-volume (<0.5 cm^3), and represent an incidental finding rather than clinically relevant disease [1,2,27]. Critically, multiclonal tumors may either be topographically distinct (multifocal), which represents the second scenario, or confluent (collision tumors), which represents the third scenario (**Figures 2 and 6**). We observed collision of clonally independent tumors, confirmed by discordant ERG IHC status, in three RP specimens (Patients 6, 11, and 12). This phenomenon of collision tumors suggests that topography alone is not reliable in determining relatedness of PCa areas.

Lastly, the disease may have a common clonal origin, but discontinuous topography (monoclonal-multifocal), potentially due to a particular form of tumor dissemination or to RP specimen processing errors; however, contrarily to what has been proposed by some authors [28], this appears to be a rare scenario (one patient in our study).

Overall, some degree of molecular heterogeneity of PCa (either in a “subclonal” or “multiclonal” form) was observed in all cases in our cohort. This suggests that random sampling of a single primary tumor area may not be sufficient to capture the entire genomic landscape of a patient’s primary disease [5].

In our study, the index tumor focus nominated by pathology in RP specimens was concordant with the one nominated by genomics in most cases where a conclusion could be reached. The three instances of discordance (Patients 2, 3, and 4) pertained to monoclonal and topographically continuous tumor areas (**supplementary material, Figures S2–S4**), which would have been considered as a unique tumor in routine pathology practice.

In contrast to our findings, Haffner *et al* reported a case in which an organ-confined tumor area with Gleason pattern 3, and not the areas with predominant pattern 4, was found to be most closely related to the metastasis [10]. However, this was not an independent low-volume Gleason grade 6 tumor, but an area within a larger lesion showing various proportions of patterns 3 and 4. In addition, this specific area also harbored a *TP53* mutation.

Comparative analysis of primary PCa and matched metastases

Paired analysis of multiple primary samples and matched metastases from individual patients is particularly helpful for determining at which point of disease progression genomic alterations of interest appear. Activating *AR* alterations were found only in metastatic samples, in agreement with the previously reported major enrichment of such alterations in metastatic cohorts [9,11,29–33], although exceptional instances of *AR* alterations in primary disease have been reported [9]. This is likely due to the fact that metastatic samples from our study represented castration-resistant disease with AR alterations driving treatment resistance [11,32]. The metastatic NEPC in Patients 1 and 7 did not show *AR* amplification, consistent with "indifference" of NEPC to AR signaling [31].

In addition, we were able to demonstrate in individual patients that some alterations reported to be enriched in metastatic cohorts readily arise in the primary disease. In particular, *TP53* alterations, which are more frequent in metastatic PCa cohorts compared to primary PCa cohorts [11,32–34], were found in metastatic samples from three patients in our study. In all three patients, these alterations were also detected in the primary cancer, and were subclonal in two cases. This could warrant further assessment of *TP53* alterations as a predictive factor for metastatic relapse in patients with localized PCa, and appears consistent with the case reported by Haffner *et al* [10]. In addition, for Patient 1, one primary tumor area (T2) and the brain metastases showed abnormal staining for P53, which usually indicates an underlying mutation (**supplementary material, Figure S1**). However, although all samples from this patient displayed a hemizygous *TP53* loss, no point mutations in *TP53* were detected by targeted DNA sequencing or by WES, despite adequate coverage. Nevertheless, we cannot exclude that accumulation of the P53 protein could be a result of alterations in other pathways that interact with P53, or of a cryptic deleterious variant in *TP53*.

Conversely to subclonal *TP53* mutations, some alterations seem to represent early ("truncal") oncogenic driver events, as they were clonally present throughout all samples from the primary disease and also shared by the metastases: *SPOP* p.F133V (a well described hotspot mutation in PCa [35,36]), *BRAF* p.K601E (a potentially activating variant, which has previously been reported in PCa [21]), and *TMPRSS2:ERG* or *ETV1* rearrangements, in keeping with the early

and clonal nature of this alteration [37]. Although to date, there are no therapies targeting these specific alterations, these examples suggest that in cases where the genomic assessment of the metastatic disease is not possible, sequencing of the primary tumor may still provide valuable information on driver mutations.

Among other recurrent alterations in PCa, *FOXA1* mutations have been reported at a frequency of 8–9% in primary disease and 12–13% in metastatic disease [38]. In our cohort, only one sample – one of the two metastases in Patient 3 – carried a *FOXA1* heterozygous p.P297 frameshift. This alteration could have an activating effect on FOXA1, as it truncates the C-terminal regulatory domain [38,39]; interestingly, this specific type of *FOXA1* alterations has been reported as significantly enriched in mCRPC compared to primary PCa [38].

Our results also highlight the potential role of epigenetic processes in metastatic PCa progression. Multiple studies have found alterations in *KMT2C* and *KMT2D*, which encode histone methyltransferases, to be enriched in metastatic PCa cohorts as compared to primary PCa [9,11,32,33,40]. Similar to *TP53*, through paired analysis, we showed that truncating *KMT2C* and *KMT2D* alterations detected in the metastases (Patients 2 and 9, respectively) were already present in the matched primaries. It could be important to investigate in larger cohorts whether the presence of such alterations in primary PCa is a risk factor for metastatic recurrence. Alterations in *KDM4A* and *ARID1A* could also warrant further investigation [21,32,40].

The frequency of pathogenic *BRCA2* germline alterations in mCRPC patients has previously been reported at 5.3% [41]. In our cohort, germline *BRCA2* alterations were found in two patients (16.7%); in one patient (Patient 1), the primary tumor and the metastases were NEPC.

This pilot study presents several limitations. Because of the limited size of this pilot cohort, findings from these “n-of-1” cases need to be confirmed in larger studies. Nevertheless, collecting archival RP specimens from patients with mCRPC is exceedingly difficult, as the period of time between primary therapy and metastatic disease can be very long (up to 18 years in our cohort, Patient 8). In addition, single cell approaches are currently the gold standard for exploring intra-tumor heterogeneity and clonal evolution, especially since some

mutations may be present in the primary tumor at low allele frequencies [42]. While this cohort of archival FFPE specimens was not suitable for single cell approaches, they should be considered in future prospective studies. Such sensitive methods would also allow to detect rare clones that were readily present in the primary tumor and selected under treatment pressure [43]. Another important hurdle in analyzing genomic relatedness between multiple samples are variations in tumor DNA content. Because of this caveat, copy number alterations could not be included in the phylogeny analyses, and SNV-based analyses may have been negatively impacted for some samples (e.g. regional lymph node samples) in which tumor purity was typically low. Further, although a recent study on multiple FFPE tissue biopsies from active surveillance PCa patients had shown consistent results between matched WES and IHC [44], the ability to compare genomic aberrations from frozen and from FFPE tissue-extracted DNA could be impaired by FFPE-related artifacts [45,46], therefore introducing an additional source of uncertainty. Lastly, other IHC biomarkers previously used in assessing PCa heterogeneity, such as PTEN and SPINK1 [47], could potentially be employed in future studies to even better discriminate tumor populations on pathology review. In particular, multiple studies have shown PTEN loss in PCa to be associated with recurrence after prostatectomy, metastatic progression and/or death [48]. In addition, the Gleason pattern 3 tumor area which proved to be the “index” focus in the case report by Haffner et al. displayed PTEN loss [10]. Thus, PTEN IHC should be employed in future study protocols of PCa heterogeneity.

Conclusion

Our study confirms that primary PCa is characterized by marked pathologic and genomic heterogeneity, which must be considered when selecting tumor areas for precision medicine studies. This is one of the first studies to formally address the still controversial concept of the “index lesion” in primary PCa. Our findings in this pilot cohort show high concordance between a comprehensive, IHC-assisted pathology review and genomic analyses for nominating the “index focus”, although the specificity of these approaches should be validated in larger cohorts.

Acknowledgements

We thank patients and their families for participating in genomics, transcriptomics, and precision cancer care studies. We are thankful for expert assistance from the Translational Research Program at WCM Pathology and Laboratory Medicine (Bing He, Leticia Dizon, Yifang Liu, Mai Ho) and the WCM CLC Genomics Core Facility (Jenny Xiang). We thank Drs. Samaneh Motanagh (Weill Cornell Medicine) and Wassim Abida (Memorial Sloan Kettering Cancer Center) for their help with retrieving clinical information for some patients. We are grateful to Makayla DeJong (Seattle Cancer Care Alliance) for her help with retrieving archival pathology material. We thank Dr. Christopher Barbieri (Weill Cornell Medicine) for his valuable suggestions regarding this study. We acknowledge expert assistance from Mariana Ricca (University of Bern) in preparing the manuscript for submission. This study was supported by the 2015-2017 Movember Foundation-PCF Challenge Award. It has also received funding from the NIH/NCI WCM SPORE in Prostate Cancer P50-CA211024 (H.B., F.D., M.A.R., J.M.M., B.D.R.), the Nuovo-Soldati Foundation (J.C.), the National Science Foundation Project grants number R37 CA241486 (H.B.), and R01 CA125612 (F.D.), the Pacific Northwest Cancer Spore Award P50CA097186 and the CDMRP Award W81XWH-18-1-0406 (P.N.), and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 648670) (F.D.).

Data availability statement

RNA-sequencing data were generated using a custom targeted amplicon-based panel and as such, do not represent transcriptome-wide gene expression profiles and were not deposited in a public repository. However, they can be made available to researchers upon request.
WES data are being deposited to the European Genome-Phenome Archive (EGA). In the meantime, processed data are available in Supplementary tables, and raw data can be shared upon request by contacting A.S. (ans2077@med.cornell.edu) or F.D. (f.demichelis@unitn.it).

Figure legends

Figure 1. A summary of all analyzed samples, including their histopathology characteristics, a schematic representation of selected genomic findings, immunohistochemistry (IHC) results, and clonal relatedness as interpreted by pathology and by genomics, i.e. phylogenetic reconstruction ("Phylogeny") and clonality score ("Clon.score"); 7+ indicates Gleason score 7 with tertiary 5 pattern; *T2:ERG* stands for *TMPRSS2:ERG*, *T2:ETV1* for *TMPRSS2:ETV1.* Selected genomic alterations are shown. To account for tumor content variability, the genomic annotation of figure panel genes was manually curated.

Figure 2. An example of pathology annotation in a case of multiclonal prostate cancer (Patient 12), including collision tumors. Pathology review and immunohistochemistry (IHC) identified 6 potentially independent tumors: a large, Gleason grade 7 focus which was ERG-/p53+ by IHC (samples T1 and T2); a Gleason grade 7 focus, which was ERG+/p53- by IHC (sample T3) and in collision with T1; a topographically distinct Gleason grade 7 focus which was ERG-/p53- by IHC (sample T5); and 3 additional low-volume distinct Gleason grade 6 foci (sample T4, pictured, and samples T6 and T7, not shown). Topographical confluence (tumor collision) was noted between areas T1 and T3, and between areas T2 and T4. The low-power view at the top is a montage of low-power views from two distinct pathology slides, taken at the same level and thus representative of one complete section of the specimen. The white dashed line indicates where the two images have been assembled.

Figure 3. A summary of pathology features, phylogeny and clonality analyses for Patient 6. (A) Pathology features of each sample, including histomorphology, ERG and p53 immunohistochemistry (IHC). Tumor focus 1 (T1) is the dominant (index) focus with highest Gleason grade and size. The T4 focus has a different ERG IHC status from other foci, and was thus regarded as an independent tumor on pathology evaluation. Aberrant p53 immunostaining is shared by T1, seminal vesicle (SV), lymph node (LN) and metastasis (Met). H&E: hematoxylin-eosin. This low-power view is a montage of low-power views from four distinct

pathology slides, taken at the same level and thus representative of one complete section of the specimen. The white dashed line indicates where the images have been assembled. (B) Phylogenetic reconstruction using 79 genes and 110 events confirms that samples T1, SV and LN are more closely related to the metastasis than samples T2 or T3, as evidenced by a common “trunk” which includes the *TP53* mutation. (C) A summary of clonality score results. Sample T1 shows the highest clonality score (highlighted in red) with respect to the metastatic sample.

Figure 4. A summary of pathology features, phylogeny and clonality analyses for Patient 9. (A) Low power pathology view. This case was interpreted as a unifocal, monoclonal tumor, albeit with the T1 tumor area showing a lower Gleason grade than areas T2 and T3. This low-power view is a montage of low-power views from two distinct pathology slides, taken at the same level and thus representing one complete section of the specimen. The white dashed line indicates where the two images have been assembled. (B) Phylogenetic reconstruction using 71 genes and 103 events, confirming that samples T2, T3 and seminal vesicle (SV) are more closely related to the metastasis than T1. Detailed histomorphology for all foci is also shown. (C) A summary of clonality score results. Sample SV shows the highest clonality score (highlighted in red) with respect to both metastatic samples. Note that sample T1 (with the lowest individual Gleason score) also shows the lowest clonality score (highlighted in blue).

Figure 5. Unsupervised sample clustering and heatmap representation of RNA-seq gene expression levels using 204 target expression amplicons. Only samples that have passed quality control as detailed in Salami *et al* [5] are shown. T2:ERG: *TMPRSS2:ERG* fusion transcripts, LN: regional lymph node, NEPC: neuroendocrine prostatic carcinoma, mCRPC: metastatic castration-resistant prostatic carcinoma. The black box indicates mCRPC samples from four patients.

Figure 6. Possible scenarios for clonality interpretation of primary prostate cancer.

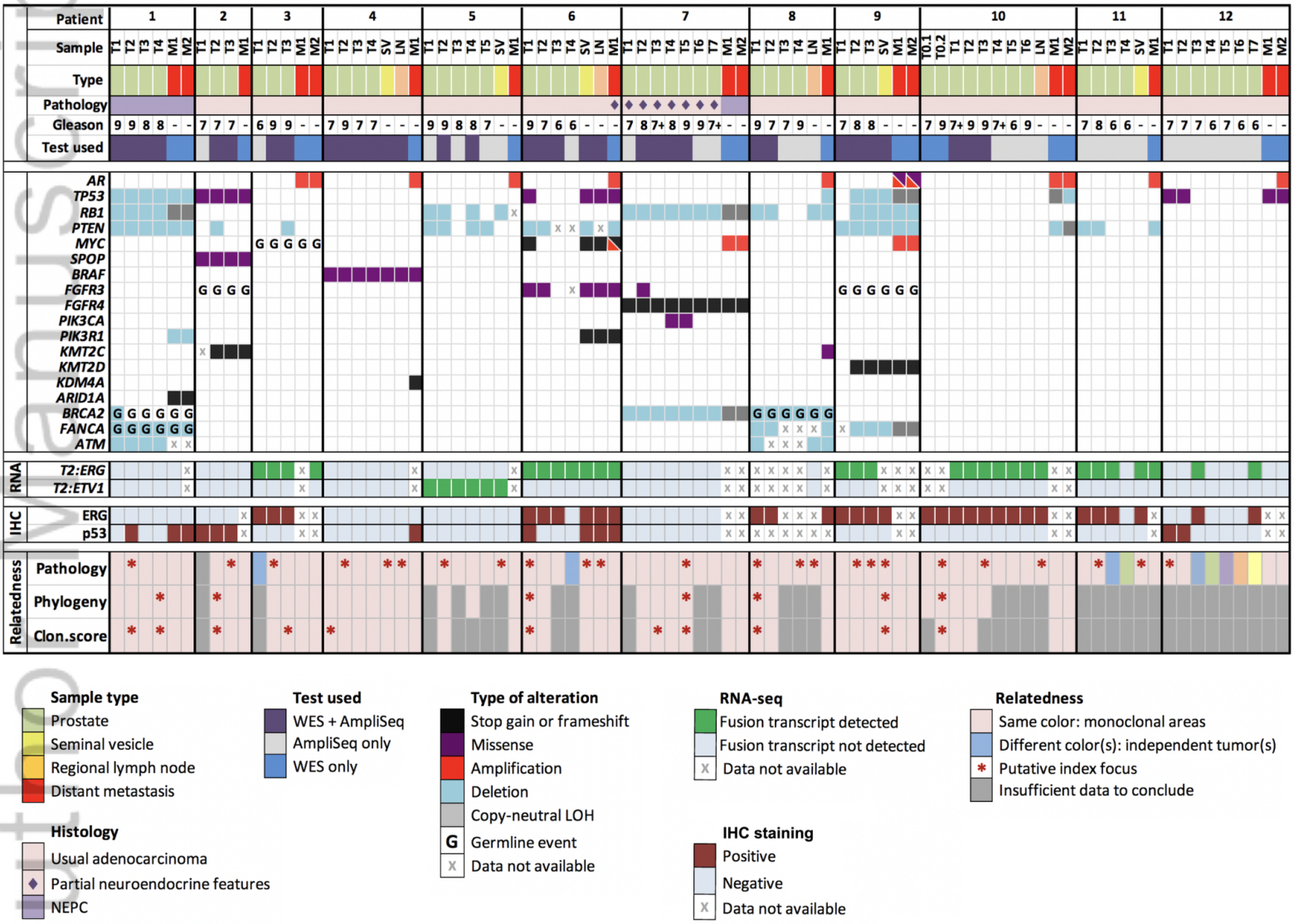
References

1. Villers A, McNeal JE, Freiha FS*, et al.* Multiple cancers in the prostate. Morphologic features of clinically recognized versus incidental tumors. *Cancer* 1992; **70**: 2313-2318.
2. Wise AM, Stamey TA, McNeal JE*, et al.* Morphologic and clinical significance of multifocal prostate cancers in radical prostatectomy specimens. *Urology* 2002; **60**: 264-269.
3. Espiritu SMG, Liu LY, Rubanova Y*, et al.* The Evolutionary Landscape of Localized Prostate Cancers Drives Clinical Aggression. *Cell* 2018; **173**: 1003-1013 e1015.
4. Rubin MA, Demichelis F. The Genomics of Prostate Cancer: emerging understanding with technologic advances. *Mod Pathol* 2018; **31**: S1-11.
5. Salami SS, Hovelson DH, Kaplan JB*, et al.* Transcriptomic heterogeneity in multifocal prostate cancer. *JCI Insight* 2018; **3**.
6. Brocks D, Assenov Y, Minner S*, et al.* Intratumor DNA methylation heterogeneity reflects clonal evolution in aggressive prostate cancer. *Cell Rep* 2014; **8**: 798-806.
7. Woodcock DJ, Riabchenko E, Taavitsainen S*, et al.* Prostate cancer evolution from multilineage primary to single lineage metastases with implications for liquid biopsy. *Nat Commun* 2020; **11**: 5070.
8. Schmidt L, Moller M, Haldrup C*, et al.* Exploring the transcriptome of hormone-naive multifocal prostate cancer and matched lymph node metastases. *Br J Cancer* 2018; **119**: 1527-1537.
9. Gundem G, Van Loo P, Kremeyer B*, et al.* The evolutionary history of lethal metastatic prostate cancer. *Nature* 2015; **520**: 353-357.
10. Haffner MC, Mosbruger T, Esopi DM*, et al.* Tracking the clonal origin of lethal prostate cancer. *J Clin Invest* 2013; **123**: 4918-4922.
11. Robinson D, Van Allen EM, Wu YM*, et al.* Integrative Clinical Genomics of Advanced Prostate Cancer. *Cell* 2015; **162**: 454.
12. Abida W, Cyrta J, Heller G*, et al.* Genomic correlates of clinical outcome in advanced prostate cancer. *Proc Natl Acad Sci U S A* 2019; **116**: 11428-11436.
13. Beltran H, Eng K, Mosquera JM*, et al.* Whole-Exome Sequencing of Metastatic Cancer and Biomarkers of Treatment Response. *JAMA Oncol* 2015; **1**: 466-474.
14. Rennert H, Eng K, Zhang T*, et al.* Development and validation of a whole-exome sequencing test for simultaneous detection of point mutations, indels and copy-number alterations for precision cancer care. *NPJ Genom Med* 2016; **1**.
15. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009; **25**: 1754-1760.
16. Cibulskis K, Lawrence MS, Carter SL*, et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* 2013; **31**: 213-219.
17. Jiang Y, Soong TD, Wang L*, et al.* Genome-wide detection of genes targeted by non-Ig somatic hypermutation in lymphoma. *PLoS One* 2012; **7**: e40332.
18. Romanel A, Lago S, Prandi D*, et al.* ASEQ: fast allele-specific studies from next-generation sequencing data. *BMC Med Genomics* 2015; **8**: 9.
19. Schliep KP. phangorn: phylogenetic analysis in R. *Bioinformatics* 2011; **27**: 592-593.
20. Ostrovnaya I, Seshan VE, Begg CB. Using Somatic Mutation Data to Test Tumors for Clonal Relatedness. *Ann Appl Stat* 2015; **9**: 1533-1548.
21. Cancer Genome Atlas Research N. The Molecular Taxonomy of Primary Prostate Cancer. *Cell* 2015; **163**: 1011-1025.
22. Epstein JI, Amin MB, Beltran H*, et al.* Proposed morphologic classification of prostate cancer with neuroendocrine differentiation. *Am J Surg Pathol* 2014; **38**: 756-767.

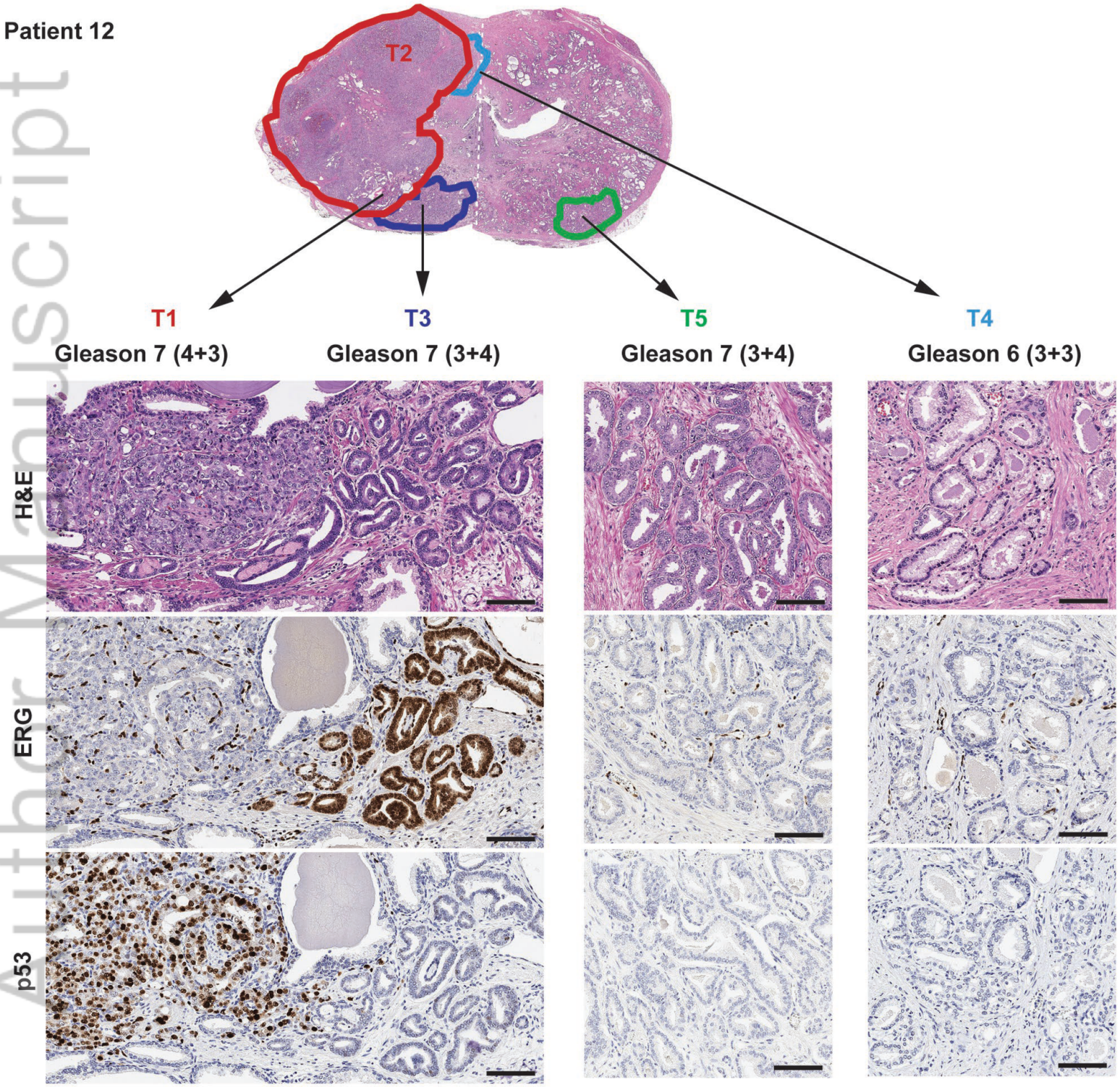
23. Cooper CS, Eeles R, Wedge DC*, et al.* Analysis of the genetic phylogeny of multifocal prostate cancer identifies multiple independent clonal expansions in neoplastic and morphologically normal prostate tissue. *Nat Genet* 2015; **47**: 367-372.
24. Lindberg J, Klevebring D, Liu W*, et al.* Exome sequencing of prostate cancer supports the hypothesis of independent tumour origins. *Eur Urol* 2013; **63**: 347-353.
25. Cheng L, Song SY, Pretlow TG*, et al.* Evidence of independent origin of multiple tumors from patients with prostate cancer. *J Natl Cancer Inst* 1998; **90**: 233-237.
26. Qian J, Bostwick DG, Takahashi S*, et al.* Chromosomal anomalies in prostatic intraepithelial neoplasia and carcinoma detected by fluorescence in situ hybridization. *Cancer Res* 1995; **55**: 5408-5414.
27. Karavitakis M, Ahmed HU, Abel PD*, et al.* Anatomically versus biologically unifocal prostate cancer: a pathological evaluation in the context of focal therapy. *Ther Adv Urol* 2012; **4**: 155-160.
28. Boyd LK, Mao X, Xue L*, et al.* High-resolution genome-wide copy-number analysis suggests a monoclonal origin of multifocal prostate cancer. *Genes Chromosomes Cancer* 2012; **51**: 579-589.
29. Liu W, Laitinen S, Khan S*, et al.* Copy number analysis indicates monoclonal origin of lethal metastatic prostate cancer. *Nat Med* 2009; **15**: 559-565.
30. Gundem G, Van Loo P, Kremeyer B*, et al.* The evolutionary history of lethal metastatic prostate cancer. *Nature* 2015; **520**: 353-357.
31. Beltran H, Prandi D, Mosquera JM*, et al.* Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer. *Nat Med* 2016; **22**: 298-305.
32. Abida W, Armenia J, Gopalan A*, et al.* Prospective Genomic Profiling of Prostate Cancer Across Disease States Reveals Germline and Somatic Alterations That May Affect Clinical Decision Making. *JCO Precis Oncol* 2017; **2017**.
33. Wedge DC, Gundem G, Mitchell T*, et al.* Sequencing of prostate cancers identifies new cancer genes, routes of progression and drug targets. *Nat Genet* 2018; **50**: 682-692.
34. Hong MK, Macintyre G, Wedge DC*, et al.* Tracking the origins and drivers of subclonal metastatic expansion in prostate cancer. *Nat Commun* 2015; **6**: 6605.
35. Barbieri CE, Baca SC, Lawrence MS*, et al.* Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. *Nat Genet* 2012; **44**: 685-689.
36. Baca SC, Prandi D, Lawrence MS*, et al.* Punctuated evolution of prostate cancer genomes. *Cell* 2013; **153**: 666-677.
37. Perner S, Mosquera JM, Demichelis F*, et al.* TMPRSS2-ERG fusion prostate cancer: an early molecular event associated with invasion. *Am J Surg Pathol* 2007; **31**: 882-888.
38. Parolia A, Cieslik M, Chu SC*, et al.* Distinct structural classes of activating FOXA1 alterations in advanced prostate cancer. *Nature* 2019; **571**: 413-418.
39. Annala M, Taavitsainen S, Vandekerkhove G*, et al.* Frequent mutation of the FOXA1 untranslated region in prostate cancer. *Commun Biol* 2018; **1**: 122.
40. Armenia J, Wankowicz SAM, Liu D*, et al.* The long tail of oncogenic drivers in prostate cancer. *Nat Genet* 2018; **50**: 645-651.
41. Pritchard CC, Mateo J, Walsh MF*, et al.* Inherited DNA-Repair Gene Mutations in Men with Metastatic Prostate Cancer. *N Engl J Med* 2016; **375**: 443-453.
42. Wang Y, Waters J, Leung ML*, et al.* Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature* 2014; **512**: 155-160.
43. Kim C, Gao R, Sei E*, et al.* Chemoresistance Evolution in Triple-Negative Breast Cancer Delineated by Single-Cell Sequencing. *Cell* 2018; **173**: 879-893 e813.
44. Gandellini P, Casiraghi N, Rancati T*, et al.* Core Biopsies from Prostate Cancer Patients in Active Surveillance Protocols Harbor PTEN and MYC Alterations. *Eur Urol Oncol* 2019; **2**: 277-285.

45. Oh E, Choi YL, Kwon MJ*, et al.* Comparison of Accuracy of Whole-Exome Sequencing with Formalin-Fixed Paraffin-Embedded and Fresh Frozen Tissue Samples. *PLoS One* 2015; **10**: e0144162.
46. Haile S, Corbett RD, Bilobram S*, et al.* Sources of erroneous sequences and artifact chimeric reads in next generation sequencing of genomic DNA from formalin-fixed paraffin-embedded samples. *Nucleic Acids Res* 2019; **47**: e12.
47. Fontugne J, Davis K, Palanisamy N*, et al.* Clonal evaluation of prostate cancer foci in biopsies with discontinuous tumor involvement by dual ERG/SPINK1 immunohistochemistry. *Mod Pathol* 2016; **29**: 157-165.
48. Jamaspishvili T, Berman DM, Ross AE*, et al.* Clinical implications of PTEN loss in prostate cancer. *Nat Rev Urol* 2018; **15**: 222-234.

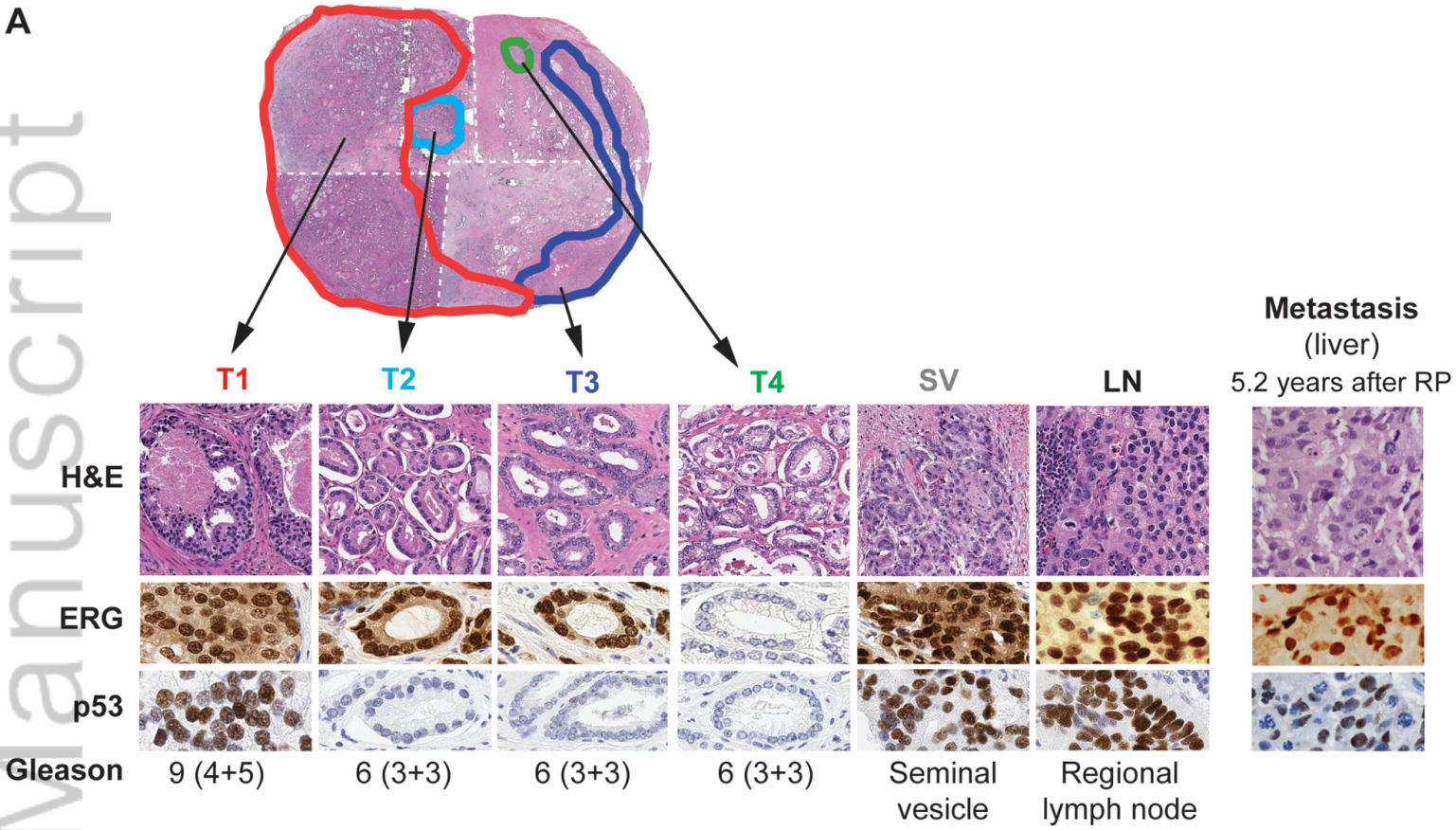
SUPPLEMENTARY MATERIAL ONLINE

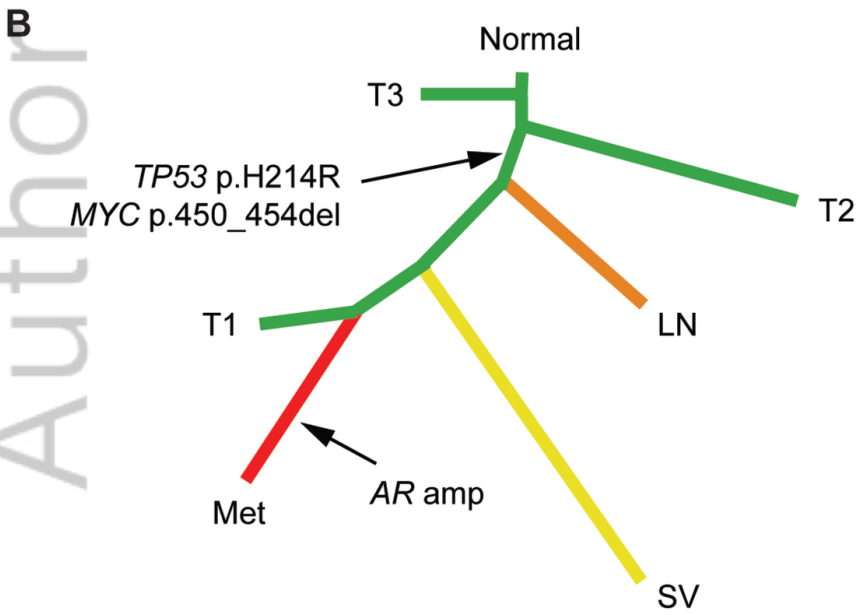


PATH_5887_Figures_v3-1_Rev_RP.png



PATH_5887_Figures_v3-2_RP.png



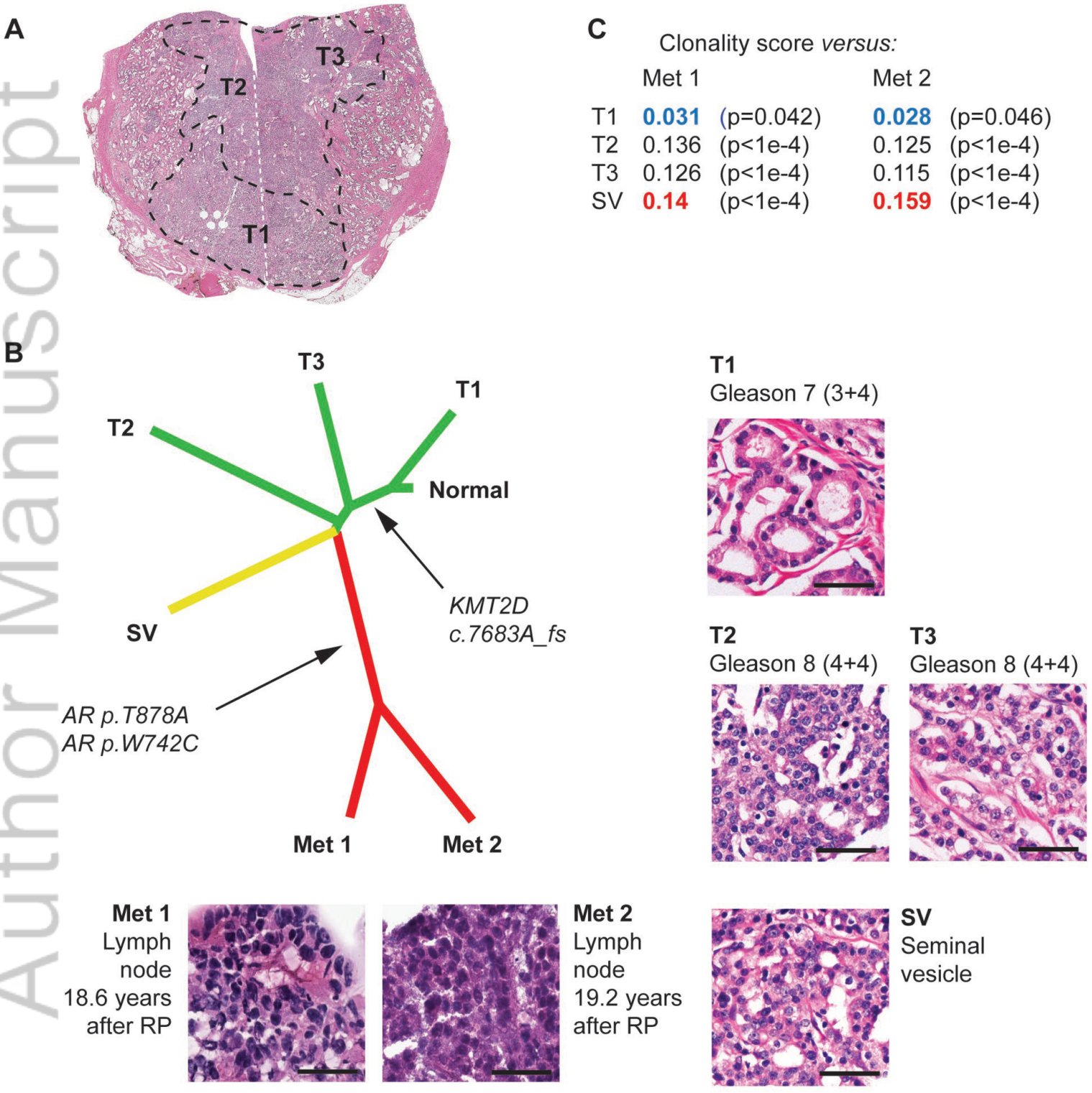


C

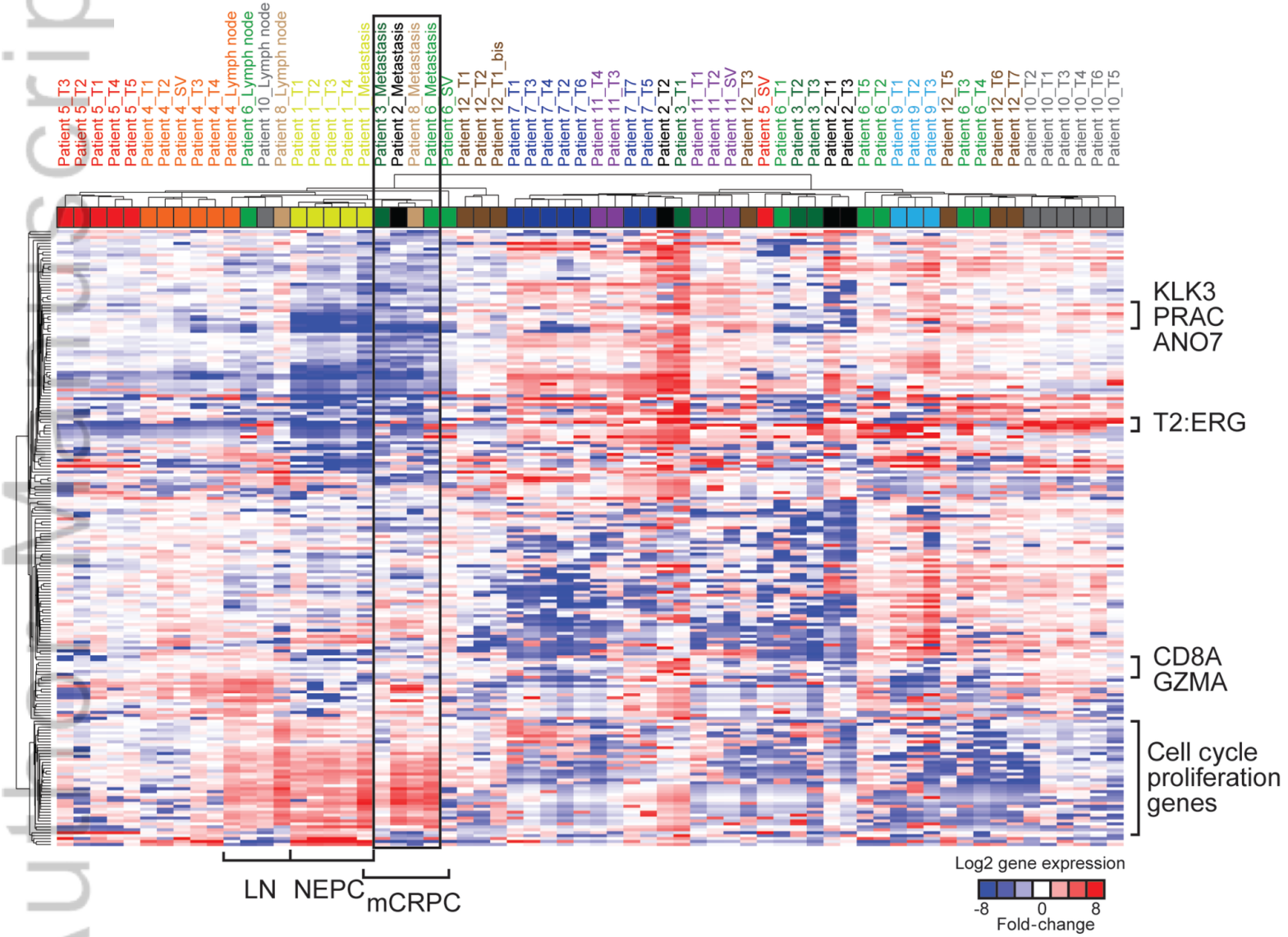
Clonality score *versus* Met

T1	**0.38**	(p<1e-4)
T2	0.09	(p<1e-4)
T3	NA	
SV	0.133	(p<1e-4)
LN	0.1	(p<1e-4)

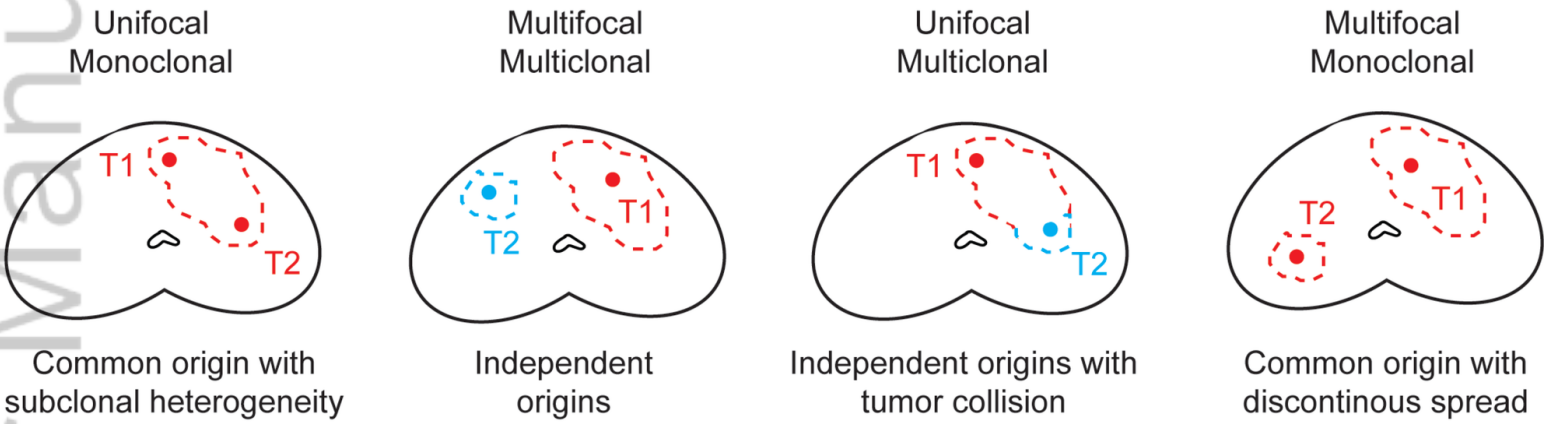
PATH_5887_Figures_v3-3_RP.png



PATH_5887_Figures_v3-4_RP.png



PATH_5887_Figures_v3-5_RP.png



PATH_5887_Figures_v3-6_RP.png